Comment on: ChatGPT: Chasing the Storm in Radiology Training and Education

Pradosh Kumar Sarangi[1] Suvrankar Datta[2] Himel Mondal[3]

[1] Department of Radiodiagnosis, All India Institute of Medical Sciences, Deoghar, Jharkhand, India
[2] Department of Radiodiagnosis, All India Institute of Medical Sciences, New Delhi, India
[3] Department of Physiology, All India Institute of Medical Sciences, Deoghar, Jharkhand, India

Indian J Radiol Imaging

Address for correspondence Pradosh Kumar Sarangi, MD, PDF, EDiR, Department of Radiodiagnosis, All India Institute of Medical Sciences, Deoghar 814152, Jharkhand, India (e-mail: drpksarangi@gmail.com).

We carefully reviewed the editorial penned by Sodhi et al, addressing the potential application of Chat Generative Pre-Trained Transformer (ChatGPT) in training and educating radiology residents, coupled with a prudent acknowledgment of its limitations and pitfalls.[1] In response, we aim to contribute some perspectives on large language models (LLMs) in radiology, from our experience of utilizing and evaluating various open-source and proprietary LLMs for various tasks in radiology.

In their editorial, the authors exclusively leveraged ChatGPT, a widely recognized and freely available LLM. It is noteworthy that while the free version of ChatGPT utilizes the GPT-3.5 model, the paid version, ChatGPT Plus and Team, leverages the GPT-4 model that boasts three key features: enhanced creativity, compatibility with visual input, and an ability to process longer contexts. GPT-4 offers unparalleled capabilities in tasks demanding advanced reasoning and complex instruction comprehension[2] and also has the ability to retrieve information from the Internet real time using retrieval augmented generation.[3]

Drawing from our personal experiences, Microsoft Bing Chat, now rebranded as Copilot, emerges as a viable alternative, allowing the utilization of the advanced GPT-4 model, without any charge. Additionally, we have explored other free LLMs such as Google Bard (rebranded as Gemini), Perplexity, and Claude.[4–6]

It is crucial to highlight that, while Gemini offers image analysis capabilities, it falls short in the domain of medical image interpretation. When prompted to interpret medical images, the response generated is shown in ►**Fig. 1**. Also recently, major issues in Gemini with respect to being biased toward certain sections of the society have raised concerns regarding the introduction of bias during the use of LLMs in sensitive domains like healthcare.

Addressing another limitation outlined by the authors, we concur with their observation regarding artificially generated references in ChatGPT responses that are often nonexistent despite exhaustive searches—a phenomenon termed "hallucination."[7] While extensive research is ongoing to figure out ways to mitigate hallucinations, we are still far from perfect.

Since the publication of the aforementioned article, there has been significant advancements in artificial intelligence (AI) capabilities for image interpretation through vision language models, custom GPTs utilizing custom prompts tailored for radiology, and more advanced methods of prompt engineering, offering potential assistance to radiologists, especially while in training.[8]

Most of the openly available LLMs currently do not have the capability of generating realistic medical images like X-rays or sections of computed tomography scans. A simple prompt action by a radiologist to show a right upper lobe consolidation on a chest X-ray does not yield the results because of guardrails set by openly available LLMs (►**Fig. 2**). Guardrails are safety controls that oversee user interaction with an LLM application, acting as rule-based systems between users and foundational models to ensure adherence to organizational principles. This is where traditional search engines are still useful, as they can access and index a vast array of medical images and information, offering immediate, although not always contextually interpreted, results. However, the scenario is changing with the introduction of

Gemini ▾

P You are an expert Radiologist. Interpret the X- ray image of the femur.

Show drafts ⌄ 🔊

✦ Sorry I can't help with that image. Try uploading another image or describing the image you tried to upload and I can help you that way.
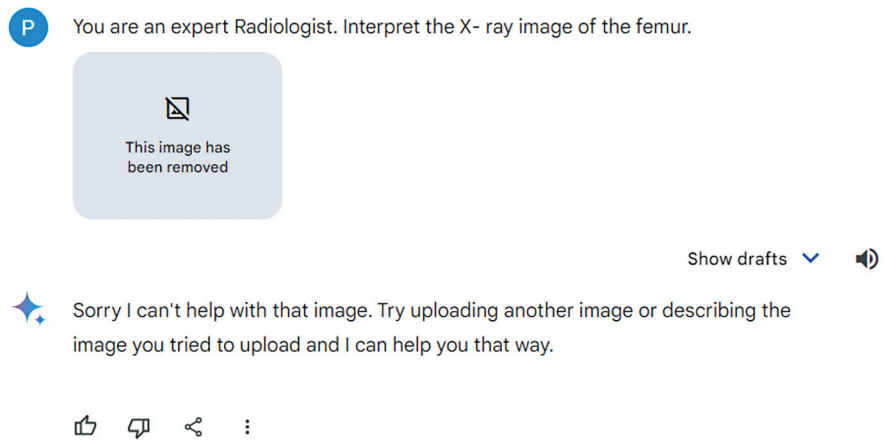
👍 👎 ⤴ ⋮

Fig. 1 Screenshot of response of Google Bard (Gemini) for the interpretation of X-ray image of the femur (response date 25-02-2024).

ChatGPT 4 ⌄

R You
You are a Radiologist's Assistant. Show me example of Right Upper Lobe Pneumonia on ChestXray

ChatGPT
I can't show actual medical images like chest X-rays directly due to privacy and legal restrictions. However, I can describe what right upper lobe pneumonia might look like on a chest X-ray.

In a chest X-ray image of right upper lobe pneumonia, you would typically see:

- An area of increased opacity or whiteness in the right upper part of the lung. This is due to the presence of infection in the lung tissue, which makes it appear denser on X-ray.
- The margins of the affected area may be ill-defined, blending into the surrounding lung tissue.
- Sometimes, the air bronchogram sign is visible, which is a pattern of air-filled bronchi (tubular structures) seen against a background of opaque, infected lung tissue.
- The right upper lobe may appear more prominent or larger due to the consolidation (solidification of the lung tissue due to fluid and infection).

These features help radiologists and physicians diagnose pneumonia and determine its location and severity.
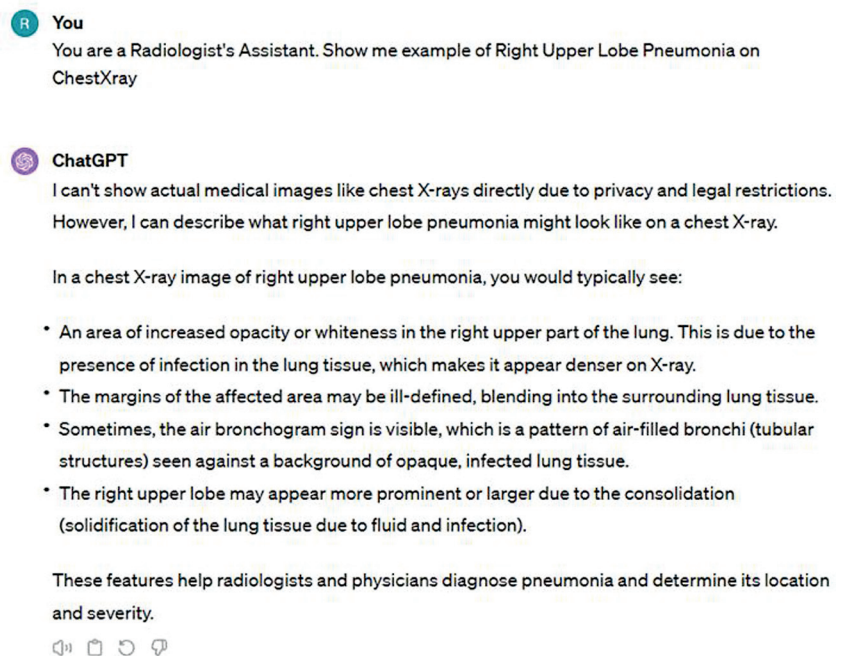
🔊 📋 ↻ 👎

Fig. 2 Screenshot of response of Chat Generative Pre-Trained Transformer 4 (ChatGPT-4) to provide image output of pneumonia on chest X-ray (response date 01-04-2024).

foundational models capable of visual question answering (VQA). VQA requires understanding natural language questions in conjunction with medical images for accurate and reliable responses. Unlike traditional search engines, these specialized VQA systems employ advanced retrieval techniques or generative capabilities to produce images directly relevant to the radiologists' queries. While general-purpose language models like GPT-4, accessible through interfaces like Copilot or ChatGPT, are evolving, radiology-specific VQA advancements are increasingly capable of responding to nuanced queries with precisely relevant images.[9,10]

In conclusion, our response aims to reflect on the original article's insights, underscore the evolving role of generative AI tools like ChatGPT in radiology, and emphasize the importance of continuous education while utilizing LLMs. We highlight the need for improvement in LLMs to mitigate current biases and hallucinations, which is pivotal for overcoming current limitations and actively utilizing LLMs in radiology.

Conflict of Interest

None declared.

References

1 Sodhi KS, Tao TY, Seymore N. ChatGPT: chasing the storm in radiology training and education. Indian J Radiol Imaging 2023;33 (04):431–435

2 GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. OpenAI. Accessed April 17, 2024 at: https://openai.com/gpt-4

3 Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv Neural Inf Process Syst 2020;33:9459–9474

4 Sarangi PK, Irodi A, Panda S, Nayak DSK, Mondal H. Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: perspectives of four large language models. Indian J Radiol Imaging 2023;34(02):269–275

5 Sarangi PK, Lumbani A, Swarup MS, et al. Assessing ChatGPT's proficiency in simplifying radiological reports for healthcare professionals and patients. Cureus 2023;15(12):e50881

6 Sarangi PK, Narayan RK, Mohakud S, Vats A, Sahani D, Mondal H. Assessing the capability of ChatGPT, Google Bard, and Microsoft Bing in solving radiology case vignettes. Indian J Radiol Imaging 2023;34(02):276–282

7 Kim S, Lee CK, Kim SS. Large language models: a guide for radiologists. Korean J Radiol 2024;25(02):126–133

8 Ji J, Hou Y, Chen X, Pan Y, Xiang Y. Vision-language model for generating textual descriptions from clinical images: model development and validation study. JMIR Form Res 2024;8:e32690

9 Narayanan A, Musthyala R, Sankar R, Nistala AP, Singh P, Cirrone J. Free form medical visual question answering in radiology. arXiv preprint arXiv:2401.13081. 2024

10 Yildirim N, Richardson H, Wetscherek MT, et al. Multimodal healthcare AI: identifying and designing clinically relevant vision-language applications for radiology. arXiv preprint arXiv:2402.14252. 2024 Feb 22